Vector Spaces, Bases, and Dual Spaces

Points, Lines, Planes and Vectors:

Strictly speaking, points are not vectors; the sum of two points is not another such point but a pair of points. However, the difference between two points can be regarded as a vector, namely the *motion* (also called *displacement* or *translation*) that carries one point to the other. In other words, adding a vector to a point yields a point; it moves the first point to the second. A vector expression like $\mathbf{b} + \mu\mathbf{c}$ moves the points' origin O first by $\mathbf{b}$, and then by a scalar multiple μ of $\mathbf{c}$. As μ runs from $-\infty$ to $+\infty$, this point runs along a straight line unless $\mathbf{c} = \mathbf{o}$, the zero vector. Alternatively, the same straight line can be represented by $\mu\mathbf{c} + \mathbf{b}$ which first runs a straight line parallel to $\mathbf{c}$ through the origin, then displaces it by $\mathbf{b}$. When the intent is clear, an expression like $\mathbf{b} + \mu\mathbf{c}$, though a vector, may sometimes be called a *point* or even a *line* if brevity is preferred over strict correctness. (Many a text-book calls the equation " $\mathbf{x} = \mathbf{b} + \mu\mathbf{c}$ " the *parametric* equation of the straight line; μ is the parameter and $\mathbf{x}$ is the displacement from origin O to a varying point on the line.)

Unless $\mathbf{c}$ and $\mathbf{d}$ are parallel, the expression $\mathbf{b} + \mu\mathbf{c} + \lambda\mathbf{d}$ displaces the origin to a point that traces out a plane as μ and λ run independently from $-\infty$ to $+\infty$. (If μ and λ did not run *independently* the figure traced out might be some curve.) The plane is swept out by parallel copies of the straight line $\lambda\mathbf{d}$ all passing through the straight line $\mathbf{b} + \mu\mathbf{c}$. Uniformly spaced copies of those lines form a cross-hatch pattern in the plane, filling it with parallelograms.

Bases of a Vector Space:

For every nonzero space of vectors $\mathbf{x}$ there are infinitely many ways to choose a coordinate system or *Basis* $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ arranged as a 1-by-n matrix of vectors $\mathbf{b}_j$ that *span* the space and are *linearly independent*. "Span" means every $\mathbf{x}$ in the space can be expressed as $\mathbf{x} = \mathbf{B}x$ if the components $\xi_1, \xi_2, .. \xi_n$ of the column vector x are chosen appropriately; *i.e.*

$$\mathbf{x} = \mathbf{b}_1\xi_1 + \mathbf{b}_2\xi_2 + \dots + \mathbf{b}_n\xi_n$$

in accordance with the rules of matrix multiplication. Every ξ_j is determined uniquely by $\mathbf{x}$ because of the *linear independence* of the $\mathbf{b}_j$'s, which means any of the following assertions:

If $\mathbf{B}x = \mathbf{o}$ then $x = o$.
If $\mathbf{B}x = \mathbf{B}y$ then $x = y$.
No basis vector $\mathbf{b}_j$ is a linear combination of the others.

Any one assertion implies the others. (Can you see why?) The integer n (assumed here to be finite) is called the *Dimension* of the vector space; it turns out to be the same no matter how the basis is chosen. To see why, suppose $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ is another basis for the same space. Spanning implies $\mathbf{C} = \mathbf{B}H$ for some n-by-m matrix H, and similarly $\mathbf{B} = \mathbf{C}K$, so $\mathbf{B}HK = \mathbf{B}$. Now linear independence implies column by column that $HK = I_n$, the n-by-n identity matrix; similarly $KH = I_m$. These equations imply $m = n$ because otherwise, say if $m > n$, a nonzero column z satisfying $Hz = o$ would have to exist (more unknowns than equations) and it would have to satisfy $o \ne z = I_m z = KHz = Ko = o$, a contradiction. Thus $I_m = I_n$ and $K = H^{-1}$; the vector space has many different bases all of the same dimension n.

What Good is a Basis ?

The freedom to choose a basis often simplifies calculations and proofs. For instance, here is a phenomenon first noticed by G. Desargues (1593 - 1662), a contemporary of R. Descartes:

> In the plane, fix two intersecting straight lines B and C and a point **p** on neither. Through **p** draw two straight lines X and Y that intersect B and C in four points all told. Two pairs of those points are not yet joined by straight lines; draw those lines now and, if they intersect, call their intersection **q** . As X and Y move, always passing through **p** , so does **q** move; show that it moves along some fixed straight line D .

To prove the existence of D takes some ingenuity if none but the methods of Euclidean plane geometry may be used; and if rectangular Cartesian coordinates must be used the proof is a tedious computation. But a relatively short computation suffices if we choose an apt basis.

> *Test these claims by trying to verify Desargues' observation above using only the ideas you learned in High-School. Then you will be better able to appreciate the strategy motivating vector notation and its algebra used in the following proof.*

Put the origin **o** at the intersection of B and C , and then choose basis vectors **b** and **c** lying along B and C respectively and with lengths so chosen that $\mathbf{p} = \mathbf{b}–\mathbf{c}$, which lies on neither B nor C . Then $\mathbf{b} + \mathbf{c}$ will turn out to lie along D , which also passes through **o** . To confirm this, let X be a line through **p** in a direction $\mathbf{b} + \xi\mathbf{c}$. Note that ξ is neither 0 nor ∞ since X is parallel to neither B nor C . Then X is traced out by $\mathbf{b}–\mathbf{c} + \mu(\mathbf{b}+\xi\mathbf{c})$ as μ runs from $–\infty$ to $+\infty$, and intersects B at $(1+1/\xi)\mathbf{b}$ when $\mu = 1/\xi$, and intersects C at $–(1+\xi)\mathbf{c}$ when $\mu = –1$. Similarly, as μ runs from $–\infty$ to $+\infty$, the line Y traced out by $\mathbf{b}–\mathbf{c} + \mu(\mathbf{b}+\eta\mathbf{c})$ for any fixed finite nonzero η intersects B at $(1+1/\eta)\mathbf{b}$ and C at $–(1+\eta)\mathbf{c}$. The last two lines to be drawn are one through $–(1+\xi)\mathbf{c}$ and $(1+1/\eta)\mathbf{b}$ traced out by

$$–(1+\xi)\mathbf{c} + \mu(\ (1+1/\eta)\mathbf{b} + (1+\xi)\mathbf{c}\ ) \quad \text{as } \mu \text{ runs from } –\infty \text{ to } +\infty ,$$

and another through $–(1+\eta)\mathbf{c}$ and $(1+1/\xi)\mathbf{b}$ traced out by

$$–(1+\eta)\mathbf{c} + \lambda(\ (1+1/\xi)\mathbf{b} + (1+\eta)\mathbf{c}\ ) \quad \text{as } \lambda \text{ runs from } –\infty \text{ to } +\infty .$$

Aptly determined values of μ and λ make these two expressions equal to a common value **q** , which is where the lines intersect. Those apt values of μ and λ are the solutions of the two linear equations obtained by equating coefficients of **b** and **c** ; the results are

$$\mu = \eta(1+\xi)/(\xi\eta – 1) ,\quad \lambda = \xi(1+\eta)/(\xi\eta – 1) , \text{ and finally}$$
$$\mathbf{q} = (\ (1+\xi)(1+\eta)/(\xi\eta – 1)\ )(\mathbf{b} + \mathbf{c}) ,$$

which runs along D as ξ and η vary with X and Y , Q. E. D.

Had a basis other than $(\mathbf{b}, \mathbf{c})$ been chosen, **b** and **c** would have been represented by some column vectors b and c other than $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$, and the point **p** would have had coordinates other than $\begin{bmatrix}1\\–1\end{bmatrix}$, and the column vector q representing **q** would have possessed components in some constant ratio more complicated than 1:1 . None the less, ***AND THIS IS IMPORTANT***, the equations relating p, b, c, q and the scalars μ, λ, ξ and η would have been the same as above except for the **bold-face**. When written out componentwise, the equations would have appeared much messier but the same algebraic manipulations would have been performed. (No such algebra was needed by Desargues, who merely imagined **o** and **p** to lie on a " line at infinity." His observation serves now as an axiom of Projective Plane Geometry, in which two lines always meet; they can't be parallel.)

Our strategy is to combine Geometry and Algebra. First we relate a geometrical or physical situation to its description in terms of abstract vectors, usually independent of a coordinate system. By choosing a convenient basis we eliminate inessential details. Then we convert to column vectors if necessary to perform the algebraic computations required. Then we convert back to the original setting. To carry out our strategy, we need geometrical interpretations for the entities and operations of vector and matrix algebra. Watch for them as we go.

Inverses of Bases, Linear Functionals, and the Dual Space

Once again let $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ be a basis for a space of vectors $\mathbf{x}$ to which column vectors are mapped by the linear operator $\mathbf{B}$. This means that for each vector $\mathbf{x}$ in the space a unique column vector x can be found to satisfy $\mathbf{x} = \mathbf{B}x$, and similarly y can be found for $\mathbf{y} = \mathbf{B}y$, and $\alpha\mathbf{x} + \beta\mathbf{y} = \mathbf{B}(\alpha x + \beta y)$. Because the "columns" of $\mathbf{B}$ are linearly independent, $\alpha x + \beta y$ is determined uniquely by the last equation, so an operator $\mathbf{B}^{-1}$ can be defined that computes the column vectors $x = \mathbf{B}^{-1}\mathbf{x}$ and $y = \mathbf{B}^{-1}\mathbf{y}$ belonging to abstract vectors $\mathbf{x}$ and $\mathbf{y}$ in the space, and it is also a linear operator: $\mathbf{B}^{-1}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha x + \beta y$. In fact, $\mathbf{B}^{-1}\mathbf{B} = I_n$, the n-by-n identity matrix; and $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$ is the identity operator that leaves every vector $\mathbf{x} = \mathbf{I}\mathbf{x}$ in the space unchanged.

Just as any square matrix can be viewed as a column of rows, so

$$\mathbf{B}^{-1} = \begin{pmatrix} \mathbf{e}^T{}_1 \\ \mathbf{e}^T{}_2 \\ \dots \\ \mathbf{e}^T{}_n \end{pmatrix}$$

is best viewed as a "column" of linear operators $\mathbf{e}^T{}_k$ that map abstract vectors to scalars; specifically, $\mathbf{e}^T{}_k\,\mathbf{x}$ is the k^{th} component ξ_k of the column vector $x = \mathbf{B}^{-1}\mathbf{x}$. Hence $\mathbf{e}^T{}_k$ is by itself the operator that, acting upon any abstract vector $\mathbf{x}$, yields the k^{th} component of the column vector x that represents $\mathbf{x}$ using the basis $\mathbf{B}$. To form a picture in two or three dimensions of what $\mathbf{e}^T{}_k$ does, draw a parallelogram or a parallelepiped with one vertex at the origin, with edges there all parallel to basis vectors $\mathbf{b}_j$, and with its opposite vertex at any chosen vector $\mathbf{x}$. Then $\mathbf{e}^T{}_k\,\mathbf{x} = \xi_k$ is the scalar multiplier that makes $\mathbf{b}_k\mathbf{e}^T{}_k\,\mathbf{x} = \mathbf{b}_k\xi_k$ the k^{th} edge of that parallelogram or parallelepiped, so $\mathbf{b}_k\mathbf{e}^T{}_k$ *projects* to this k^{th} edge in a direction parallel to the line or plane containing all other edges. *Draw pictures!*

From any row $w^T = (\omega_1, \omega_2, \dots, \omega_n)$ construct $\mathbf{w}^T = w^T\mathbf{B}^{-1} = \omega_1\mathbf{e}^T{}_1 + \omega_2\mathbf{e}^T{}_2 + \dots + \omega_n\mathbf{e}^T{}_n$, which acts upon a vector $\mathbf{x} = \mathbf{B}x$ whose column x has components $\xi_1, \xi_2, \dots, \xi_n$ to produce $\mathbf{w}^T\mathbf{x} = w^Tx = \omega_1\xi_1 + \omega_2\xi_2 + ... + \omega_n\xi_n$. This *scalar product* is so important that it has many aliases: $\mathbf{w}\bullet\mathbf{x}$, $<\mathbf{x},\mathbf{w}>$, $<\mathbf{x}|\mathbf{w}>$ and $(\mathbf{x}, \mathbf{w})$ are some of them, but we shall see later that they are not exact equivalents. The entity $\mathbf{w}^T$ is a *linear functional*, a linear operator because $\mathbf{w}^T(\mu\mathbf{x} + \lambda\mathbf{y}) = \mu\mathbf{w}^T\mathbf{x} + \lambda\mathbf{w}^T\mathbf{y}$, and a *functional* rather than a *function* because its values are exclusively scalar. A geometrical interpretation for it will be offered in a few moments.

Historical digression: The noun "functional" arose first from the adjective in *Functional Analysis*, which was at first concerned with operators that map functions to scalars; an instance is the definite integral $h(f) = \int_o^1 f(\kappa)\,d\kappa$ regarded as an operator upon functions $f(\kappa)$. In time, each function $f(\kappa)$ became identified with a point and then a vector $\mathbf{f}$ in a space of functions. Then it became possible to write $h(f) = \mathbf{h}^T\mathbf{f}$ where the linear functional $\mathbf{h}^T$ stood for the integral operator $\int_o^1 \dots(\kappa)\,d\kappa$ divorced from the function f upon which it acted. Note how the functional $\mathbf{h}^T$ is not the transpose of a vector $\mathbf{h}$; there is no vector $\mathbf{h}$. Pedants who wished to stress this non-relationship used to choose, say, $\mathbf{v}^T$ for a functional and $\mathbf{v}$ for an unrelated function, like a playwright choosing names "Edwin" and "Edwina" for unrelated characters. That's a confusing choice I intend to avoid.

Now observe that every linear functional $\mathbf{w}^T$ acting upon the given vector space, no matter how $\mathbf{w}^T$ is constructed, must satisfy $\mathbf{w}^T(\mu\mathbf{x} + \lambda\mathbf{y}) = \mu\mathbf{w}^T\mathbf{x} + \lambda\mathbf{w}^T\mathbf{y}$, and therefore must be expressible in the form $\mathbf{w}^T = w^T\mathbf{B}^{-1}$ by setting $\omega_j = \mathbf{w}^T\mathbf{b}_j$ in the row vector w^T . *Work this out!* Therefore the set of all these linear functionals $\mathbf{w}^T$ constitutes another vector space, called the *Dual* or *Conjugate* of the space of vectors $\mathbf{x}$, with the "rows" $\mathbf{e}^T{}_k$ of $\mathbf{B}^{-1}$ as a basis. The two spaces are situated symmetrically, dual to each other (provided the dimension n is finite). For instance, dual to the space of row n-vectors is the space of column n-vectors; their only difference appears now to be that one space's vectors are written on the left, the other's on the right of the scalar product. Other differences will turn up later.

To any fixed scalar ß and functional $\mathbf{w}^T \neq \mathbf{o}^T$ corresponds the locus traced out by all vectors $\mathbf{x}$ that satisfy the equation $\mathbf{w}^T\mathbf{x} = ß$. This locus divides the space of vectors $\mathbf{x}$ into two half-spaces, one where $\mathbf{w}^T\mathbf{x} > ß$ and one where $\mathbf{w}^T\mathbf{x} < ß$. To gauge the shape of this locus, we find that if it contains $\mathbf{x}_1$ and $\mathbf{x}_2 \neq \mathbf{x}_1$ then it contains the whole straight line $\mathbf{x}_1 + \mu(\mathbf{x}_2 - \mathbf{x}_1)$ through them. Therefore the locus is that straight line if the space of vectors $\mathbf{x}$ is just two-dimensional. Similarly, if the locus contains $\mathbf{x}_1$ and $\mathbf{x}_2$ and $\mathbf{x}_3$ with nonzero and non-parallel $\mathbf{x}_2\text{-}\mathbf{x}_1$ and $\mathbf{x}_3\text{-}\mathbf{x}_1$, then the whole plane $\mathbf{x}_1 + \mu(\mathbf{x}_2\text{-}\mathbf{x}_1) + \lambda(\mathbf{x}_3\text{-}\mathbf{x}_1)$ lies in the locus too. The locus is that plane if the whole space is just three-dimensional. In spaces of higher dimension the locus is a *hyperplane* whose dimension is one less than the dimension of the space. To complete this picture we should confirm that the locus really has enough points in it properly situated to generate a hyperplane; that confirmation requires a basis $\mathbf{B}$ for the space. Using this basis, the equation $\mathbf{w}^T\mathbf{x} = ß$ takes the form $w^Tx = ß$, *i.e.*

$$\omega_1\xi_1 + \omega_2\xi_2 + \ldots + \omega_n\xi_n = ß ,$$

which is one linear equation with n unknowns $\xi_1, \xi_2, \ldots, \xi_n$. Suppose $\omega_n \neq 0$; then the first n–1 of those unknowns may be chosen arbitrarily. For instance, set all of them to 0 , or set one of them to 1 and the rest to 0 , and then solve for ξ_n . This is a way to generate the needed n points $\mathbf{x}_j$ with n–1 linearly independent differences. *Try it for* n = 2 *and for* n = 3. There are other ways; what are the hyperplane's *intercepts* on the coordinate axes?

If we wish to think of a set of linear functionals as a set of "vectors" in the dual-space, how do we assign direction and magnitude to a "vector" (functional) $\mathbf{w}^T$? Its direction is the *normal* shared by the family of parallel loci $\mathbf{w}^T\mathbf{x} = ß$ generated by changing ß . And if ß is stepped through uniformly distributed values, say ß = …, –2, –1, 0, 1, 2, 3, … , then the corresponding loci form a ruling of uniformly spaced lines or planes or hyperplanes, each separated from its two neighbors by a distance that is inversely proportional to the magnitude of $\mathbf{w}^T$, as can be verified by observing that for any $\mu > 0$ the set of loci whose equations are $(\mu\mathbf{w}^T)\mathbf{x} = ß$ is the same as the set whose equations are $\mathbf{w}^T\mathbf{x} = ß/\mu$. The foregoing notions of direction and magnitude seem vague because they are so general; they do not require a way to compare magnitudes of non-parallel vectors nor to measure angles between them.

Footnote: "Duality" sometimes means a 1-to-1 map between hyperplanes in one space and points in its dual space, omitting $\mathbf{o}$ and $\mathbf{o}^T$, set up by an equation like $\mathbf{w}^T\mathbf{x} = 1$.

Change of Basis:

Suppose now that $\mathbf{B}$ and $\mathbf{C}$ are two different bases for the same space. We have seen above that $\mathbf{B} = \mathbf{C}K$ and $\mathbf{C} = \mathbf{B}K^{-1}$ for some matrix K . Some author calls it the *transition matrix* from $\mathbf{B}$ to $\mathbf{C}$; or is it from $\mathbf{C}$ to $\mathbf{B}$? I can't remember.

If the coordinates of $\mathbf{x}$ in basis $\mathbf{B}$ form a column $x = \mathbf{B}^{-1}\mathbf{x}$, then the coordinates of $\mathbf{x}$ in basis $\mathbf{C}$ must form a column $\mathbb{x} = \mathbf{C}^{-1}\mathbf{x} = K\mathbf{B}^{-1}\mathbf{x} = Kx$. And if the coordinates of $\mathbf{w}^T$ in basis $\mathbf{B}$ form a row $w^T = \mathbf{w}^T\mathbf{B}$, then the coordinates of $\mathbf{w}^T$ in basis $\mathbf{C}$ must form a row $\mathbb{w}^T = \mathbf{w}^T\mathbf{C} = \mathbf{w}^T\mathbf{B}K^{-1} = w^TK^{-1}$. Hence, any change of basis can be effected by multiplying coordinates by a suitable nonsingular (invertible) matrix or its inverse. Which? Don't try to memorize which; repeating the foregoing manipulations is easier and more reliable.

Note that the scalar product $\mathbf{w}^T\mathbf{x} = w^Tx = \mathbb{w}^T\mathbb{x}$ is the same no matter which basis is used; this is another instance of the important principle mentioned above:

> Geometrically meaningful coordinate-free equations involving scalars, abstract vectors and functionals remain unchanged in form after a basis is used to convert vectors into columns and functionals into rows, and these equations are therefore unchanged by changes of basis. The choice of basis determines numerical values for coordinates, some of which vanish conveniently if the basis is chosen for this purpose.

Many notations for vectors violate this principle because, designed initially for Euclidean vector spaces, these notations fail to distinguish a space from its dual. Failure is inevitable because there is no way to distinguish an Euclidean space from its dual. These notations appear in many a text-book and in computer software where vectors are represented sometimes by rows, sometimes by columns, thus giving rise to confusing formulas and notations like $(\mathbf{v})_B$ for the row and $[\mathbf{v}]_B$ for the column that represent the same abstract vector $\mathbf{v}$ relative to a varying basis $\mathbf{B}$. Confusion intensifies when the kind of symbol "$\mathbf{v}$" used for an abstract vector $\mathbf{v}$ is used also in one place for a row, in another for a column, so you can't tell whether expressions involving several such symbols are well-formed without scrutinizing their context. This kind of confusion is so widespread that you must learn to cope with it; but I hope you won't take that as a license to imitate other persons' mistakes.

To treat the symbol " $\mathbf{w}^T$ " in use for an abstract functional as if it were an abstract vector $\mathbf{w}$ transposed is a mistake. Although " G^T " stands for the transpose of a *matrix* G , there is no universally valid relation between an abstract functional $\mathbf{w}^T$ and an abstract vector $\mathbf{w}$ with the same name "w". The trouble is not that no such relations exist, but that there are too many of them. Choosing one, establishing a map between the vectors in a space and the functionals in its dual space, turns the spaces into something special:– real Euclidean, or complex Unitary, or Hilbert or Minkowski or Banach or … spaces.

To see how that happens suppose we try to associate real vectors $\mathbf{w}$ with functionals $\mathbf{w}^T$ by giving them equal coordinates in every coordinate system, writing those coordinates in a row w^T for $\mathbf{w}^T$, a column w for $\mathbf{w}$. When we change to a new coordinate system that now uses column vector $\mathbb{v} = Kv$ to represent abstract vector $\mathbf{v}$ formerly represented by column v , we must now represent functional $\mathbf{f}^T$, formerly represented by row f^T , by row $\mathbb{f}^T = f^TK^{-1}$ in order to keep $\mathbf{f}^T\mathbf{v} = f^Tv = (f^TK^{-1})(Kv) = \mathbb{f}^T\mathbb{v}$. As a vector, $\mathbf{w}$ formerly represented by w is now represented by $\mathbb{w} = Kw$; as a functional, $\mathbf{w}^T$ formerly represented by w^T is represented now by $\mathbb{w}^T = w^T K^{-1}$. For the new coordinates of $\mathbf{w}$ and $\mathbf{w}^T$ to be equal, as the old were assumed to be, $\mathbb{w}^T$ must be the transpose of $\mathbb{w}$, which means $w^T K^{-1} = (Kw)^T = w^T K^T$. If this equation must hold for every w^T , it implies $K^{-1} = K^T$; can you see why?

In short, to be consistent with an association between a real vector space and its dual that gives each vector $\mathbf{w}$ and associated functional $\mathbf{w}^T$ equal coordinates in every coordinate system, we must restrict coordinate transformations' matrices K to satisfy $K^T = K^{-1}$. Such matrices are very special; called *orthogonal matrices*, they represent basis changes from one *orthonormal* coordinate system to another in a *Euclidean* space whose every vector $\mathbf{v}$ has a *length* $\|\mathbf{v}\|$ defined by the familiar *Pythagorean* formula from the coordinates of $\mathbf{v}$ thus:

$$\|\mathbf{v}\| := \sqrt{(\mathbf{v}^T\mathbf{v})} = \sqrt{(v^T v)} = \sqrt{(\text{ sum of squares of coordinates of } \mathbf{v}\ )}\ .$$

It is easy to verify that this formula for length is independent of orthogonal changes of basis; *i.e.* $\mathbb{v}^T\mathbb{v} = (Kv)^T(Kv) = v^T v$ whenever K is an orthogonal matrix. More interesting is the deduction that K must be orthogonal if the last equation is satisfied for every v ; can you prove this? Later we shall see that infinitely many n-by-n orthogonal matrices exist, every one the product of at most n elementary *orthogonal reflections* each of the form $I - (2/c^Tc)cc^T$ for a different column c . Products of even numbers of reflections turn out to be *rotations*.

Euclidean, non-Euclidean and Affine Spaces

The vector spaces with which you are best acquainted are Euclidean spaces. Each is a real space with an orthonormal basis $\mathbf{B}$ in which every vector $\mathbf{v} = \mathbf{B}v$ has a length computed from its coordinates by the Pythagorean formula $\|\mathbf{v}\| = \sqrt{(v^T v)}$. Another orthonormal basis $\mathbf{C} = \mathbf{B}K^{-1}$ can be obtained by post-multiplying by any orthogonal matrix $K^{-1} = K^T$. If vector $\mathbf{w} = \mathbf{B}w$ and functional $\mathbf{w}^T = w^T\mathbf{B}^{-1}$ are related by equal (though transposed) coordinates w and w^T with one orthonormal basis, the equality of their coordinates persists with every orthonormal basis.

Thus, an Euclidean space is its own dual space. But this relationship is spoiled by the choice of a non-orthonormal basis. Then $\mathbf{w}$ and $\mathbf{w}^T$ no longer have equal coordinates, and length no longer satisfies the Pythagorean formula. If you were given a vector space but not an orthornormal basis for it, could you tell whether the space is Euclidean? It might not be.

A way to tell was found early in this century by C. Jordan and J. von Neumann. It is a test applied to the lengths of vectors, provided length is defined. (There are vector spaces for which length is undefined, or is defined differently than by the Pythagorean formula.)

Theorem: An orthonormal basis exists if and only if length satisfies the *Parallelogram Identity* :

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 \equiv 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \ ,$$

and then the Euclidean association between vectors and functionals is provided by the formula

$$\mathbf{w}^T\mathbf{x} = (\ \|\mathbf{x} + \mathbf{w}\|^2 - \|\mathbf{x} - \mathbf{w}\|^2\ )/4\ .$$

The theorem's proof is not obvious. It will be presented later in this course.

The complex vector space analogous to a real Euclidean space is a complex *Unitary* space. The most obvious differences are changes in notation from transposes w^T to complex conjugate transposes w^*, from *orthogonal* matrices satisfying $K^T = K^{-1}$ to *unitary* matrices satisfying $K^* = K^{-1}$, and from the Pythagorean formula for Euclidean length to $\|\mathbf{w}\| := \sqrt{(\mathbf{w}^*\mathbf{w})} = \sqrt{(w^*w)} = \sqrt{(}$ sum of squared magnitudes of coordinates of $\mathbf{w}$) . The Parallelogram identity stays the same, but the formula after it provides $\mathrm{Re}(\mathbf{w}^*\mathbf{x})$.

Non-Euclidean (and non-Unitary) vector spaces do exist. The simplest example is infinite-dimensional: Consider the space of columns with infinitely many *rational* components of which only finitely many can be nonzero. It is a vector space because such columns can be multiplied by rational scalars and added to get more of the same. The dual space consists of rows with infinitely many rational components chosen arbitrarily. This dual space is not like the original vector space at all; there are more rows than there are columns.

The simplest vector spaces have no special relation between a space and its dual (other than that the dual space's vectors are written on the left side of the scalar product) and no definition of length for vectors, though two parallel vectors' lengths can be compared because one is a scalar multiple of the other. These spaces are called *Affine* spaces; their geometries concerns the properties of configurations of objects that persist after Affine transformations that map points (vectors) to points, lines to lines, planes to planes, …, hyperplanes to hyperplanes, and preserve parallelism. (If parallelism is not preserved the geometry is *Projective*.) Although these spaces lack a definition for vector length, we shall see later that they do posses a definition for *Content* :– area or volume or hypervolume. Every other kind of vector space can be obtained from an affine space by specializing it in some way, usually by adding a definition for vector length or defining a map between the vector space and its dual. In the absence of such a map, and usually when one exists too, we will find that linear algebra works best when the notation distinguishes between a space and its dual.

Example: Cubic Polynomials

Cubic polynomials $p(\xi) = \pi_o + \pi_1\xi + \pi_2\xi^2 + \pi_3\xi^3$ constitute a vector space because the sum of two of them is another, and the product of one with a constant scalar is yet another. Note that $\pi_3 = 0$ is allowed. The coefficients $\pi_o, \pi_1, \pi_2, \pi_3$ are arbitrary constant scalars but ξ is an indeterminate. The cubic whose coefficients are all 0 is the zero vector in the space; it is the only cubic that takes the value 0 for all ξ because a nonzero cubic cannot vanish at more than three values ξ . (Why not?) That's why the "columns" of $\mathbf{U} := (1, \xi, \xi^2, \xi^3)$ are linearly independent functions; they constitute a basis for the space, which therefore has dimension 4 .

The usual notation for functions is ambiguous. When we see " $p(\xi) = \pi_o + \pi_1\xi + \pi_2\xi^2 + \pi_3\xi^3$ " out of context we cannot tell whether it stands for the cubic polynomial or for its value when its indeterminate takes the value ξ . Let us now agree upon the latter interpretation, and write $\mathbf{p}$ for the cubic as a whole. In a similar spirit, we write $\mathbf{u}_n$ for the polynomial whose value is $u_n(\xi) = \xi^n$. Then we see $\mathbf{U} := (\mathbf{u}_o, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ is the aforementioned basis for the space of cubics, and that $\mathbf{p} = \mathbf{U}p$ for a column vector p whose components are π_o, π_1, π_2 and π_3 . What is $\mathbf{U}^{-1}$? Somehow it extracts the components of $p = \mathbf{U}^{-1}\mathbf{p}$ from the "cubic as a whole."

Exercise: Choose a constant scalar ß and let $\mathbf{v}_n$ stand for the polynomial $v_n(\xi) := u_n(\xi + ß)$, so that $\mathbf{v}_n = \mathbf{u}_n + nß\mathbf{u}_{n-1} + n(n\text{-}1)ß^2\mathbf{u}_{n-2}/2 + \dots$. For the space of cubics, $\mathbf{V} := (\mathbf{v}_o, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ is another basis, one that corresponds to a "shifted" origin of the variable ξ ; *i.e.*, $\mathbf{V}$p is the cubic $p(\xi + ß)$ whose expansion in powers of ξ has coefficients $\mathbf{U}^{-1}\mathbf{V}$p . Exhibit matrix $\mathbf{U}^{-1}\mathbf{V}$ and its inverse explicitly in terms of ß . What does $\mathbf{V}\mathbf{U}^{-1}$ do?

What relates the cubic $\mathbf{p}$ to one of its values $p(\xi)$ is an operator $\mathbf{e}^T(\xi)$ called *evaluation at* ξ and defined by $\mathbf{e}^T(\xi)\mathbf{p} = p(\xi)$. Evidently $\mathbf{e}^T(\xi)$ is a *linear functional* because, if $\mathbf{p}$ and $\mathbf{q}$ are two cubics, $\mathbf{e}^T(\xi)(\mathbf{p} + \mathbf{q}) = p(\xi) + q(\xi) = \mathbf{e}^T(\xi)\mathbf{p} + \mathbf{e}^T(\xi)\mathbf{q}$. In the basis $\mathbf{U}$ the components of evaluation functional $\mathbf{e}^T(\xi)$ constitute a row vector $\mathrm{e}^T(\xi) = \mathbf{e}^T(\xi)\mathbf{U} = (1, \xi, \xi^2, \xi^3)$ of scalar values (not functions as they were for $\mathbf{U}$), whence $\mathbf{e}^T(\xi)\mathbf{p} = \mathrm{e}^T(\xi)\mathrm{p} = p(\xi)$ as expected.

Interpolation is the recovery of a function from samples of its values. Let $X := \{\xi_1, \xi_2, \xi_3, \xi_4\}$ be a set of four distinct values of ξ ; we shall recover the whole cubic $\mathbf{p}$ from four sample-values $\{p(\xi_1), p(\xi_2), p(\xi_3), p(\xi_4)\}$. Written as a column vector, these sample-values are the result $\mathbf{E}^T(X)\mathbf{p}$ of applying to the cubic $\mathbf{p}$ a linear operator $\mathbf{E}^T(X)$ comprising four evaluation functionals $\{\mathbf{e}^T(\xi_1), \mathbf{e}^T(\xi_2), \mathbf{e}^T(\xi_3), \mathbf{e}^T(\xi_4)\}$ written in a column. Interpolation is the inversion of $\mathbf{E}^T(X)$. To accomplish this define quartic polynomial $\varsigma(\xi, X) := (\xi-\xi_1)(\xi-\xi_2)(\xi-\xi_3)(\xi-\xi_4)$ and its derivative $\varsigma'(\xi, X) := \partial\varsigma(\xi, X)/\partial\xi$ and then four cubics $\mathbf{c}_j(X)$ whose values are $\mathbf{e}^T(\xi)\mathbf{c}_j(X) = c_j(\xi, X) := \varsigma(\xi, X)/\big((\xi-\xi_j)\varsigma'(\xi_j, X)\big)$ for j = 1, 2, 3, 4 . If these cubics are linearly independent they can be assembled into a basis $\mathbf{C}(X) := (\mathbf{c}_1(X), \mathbf{c}_2(X), \mathbf{c}_3(X), \mathbf{c}_4(X))$. Confirm that $\mathbf{C}(X)$ is a basis for the space of cubics by proving that $\mathbf{E}^T(X)\mathbf{C}(X) = \mathrm{I}$; and then confirm that $\mathbf{C}(X) = \mathbf{E}^T(X)^{-1}$ by proving that $\mathbf{C}(X)\mathbf{E}^T(X) = \mathbf{I}$, the identity on the space of cubics. (Do you see why $\mathbf{C}(X)\mathbf{E}^T(X)\mathbf{p} = \mathbf{p}$?) The operator $\mathbf{C}(X)$ is called " Lagrange Interpolation."

To every function $w(\tau)$ continuous on $-1 \le \tau \le 1$ is associated a linear functional $\mathbf{w}^T$ thus: $\mathbf{w}^T\mathbf{p} = \int_{-1}^1 w(\tau)\, p(\tau)\, d\tau$. This linear functional's representation as a row $\mathrm{w}^T = (\omega_o, \omega_1, \omega_2, \omega_3)$ in basis $\mathbf{U}$ is $\mathrm{w}^T = \mathbf{w}^T\mathbf{U}$, so $\omega_n = \mathbf{w}^T\mathbf{u}_n = \int_{-1}^1 w(\tau)\, \tau^n\, d\tau$. Can *every* linear functional $\mathbf{m}^T$ acting upon cubics $\mathbf{p}$ be represented as an integral $\mathbf{m}^T\mathbf{p} = \int_{-1}^1 m(\tau)\, p(\tau)\, d\tau$ and thus associated with some continuous function $m(\tau)$? Yes; one such $m(\tau)$ is derived from $\mathbf{m}^T$ as follows:

The expression $h(\xi, \tau) = (9 - 15\tau^2 + 75\tau\xi - 105\tau^3\xi - 15\xi^2 + 45\tau^2\xi^2 - 105\tau\xi^3 + 175\tau^3\xi^3)/8$ is a cubic polynomial in τ as well as ξ . Think of it as the cubic $\mathbf{h}(\tau) = \mathbf{U}\mathrm{h}(\tau)$ where column $\mathrm{h}(\tau)$ has components $(9-15\tau^2)/8$, $(75\tau - 105\tau^3)/8$, $(-15 + 45\tau^2)/8$ and $(-105\tau + 175\tau^3)/8$. Now $m(\tau) = \mathbf{m}^T\mathbf{h}(\tau)$ turns out to have the desired property: $\mathbf{m}^T\mathbf{p} = \int_{-1}^1 m(\tau)\, p(\tau)\, d\tau$ for *every* cubic $\mathbf{p}$, as can be verified by confirming that $\int_{-1}^1 \mathrm{h}(\tau)\, \mathrm{e}^T(\tau)\, d\tau = \mathrm{I}$ (*you must do this yourself*) and then that $\int_{-1}^1 \mathbf{h}(\tau)\, \mathbf{e}^T(\tau)\, d\tau = \mathbf{I}$ is the identity operator on the space of cubics. Because of the last two equations, the expression $h(\xi, \tau) = h(\tau, \xi) = \mathrm{e}^T(\xi)\mathrm{h}(\tau) = \mathbf{e}^T(\xi)\mathbf{h}(\tau)$ is called a *reproducing kernel*. Can you figure out how it was constructed?

Thus we see that every linear functional $\mathbf{m}^T$ acting on cubics $\mathbf{p}$ is representable either as a row vector m^T acting upon their columns p of coefficients, or as an integral acting upon their values $p(\xi)$, and in the integral $\mathbf{m}^T\mathbf{p} = \int_{-1}^1 m(\tau)\,p(\tau)\,d\tau$ the function $m(\tau)$ associated with $\mathbf{m}^T$ can be chosen in many ways, one of them as a uniquely determined cubic ! This identifies the space of cubics with its dual space in such a way as to turn them into an Euclidean space wherein length is defined by $\|\mathbf{p}\| := \sqrt{(\int_{-1}^1 p(\tau)^2\,d\tau)} = \sqrt{(p^T A p)}$ for a suitable matrix A *not* the identity matrix. (Can you determine A ? Compare A^{-1} with $h(\xi, \tau)$.) Because basis $\mathbf{U}$ is not orthonormal, the column of coefficients $\mathbf{U}^{-1}\mathbf{m}$ of the cubic polynomial $\mathbf{m}$, whose value $\mathbf{e}^T(\tau)\mathbf{m} = m(\tau)$ is determined by the functional $\mathbf{m}^T$, are not the same as the elements of the row vector m^T that delivers $\mathbf{m}^T\mathbf{p} = m^T p$ from the column of coefficients $p = \mathbf{U}^{-1}\mathbf{p}$ of every cubic $\mathbf{p}$. Instead $m^T = (\mathbf{U}^{-1}\mathbf{m})^T A$; can you see why? In short, if $\mathbf{m}$ and $\mathbf{p}$ are cubics we can evaluate $\mathbf{m}^T\mathbf{p}$ either from its definition as an integral or directly from their coefficients using the last three equations and the matrix A .

Polynomials $\mathbf{q}_o := \mathbf{u}_o/\sqrt{2}$, $\mathbf{q}_1 := \mathbf{u}_1\sqrt{(3/2)}$, $\mathbf{q}_2 := (3\mathbf{u}_2 - \mathbf{u}_o)\sqrt{(5/8)}$, $\mathbf{q}_3 := (5\mathbf{u}_3 - 3\mathbf{u}_1)\sqrt{(7/8)}$ provide an orthonormal basis $\mathbf{Q} := (\mathbf{q}_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ for the space of cubics. Prove this by verifying $\mathbf{Q}^T\mathbf{Q} = I$; you can compute integrals or matrix products to do it.

Now we can address the following question: How few samples of a cubic $\mathbf{p}$ suffice to determine its definite integral $\mathbf{i}^T\mathbf{p} := \int_{-1}^1 p(\tau)\,d\tau$? This integral defines a linear functional $\mathbf{i}^T$, and we wish to express it in terms of as few evaluations $\mathbf{e}^T(ß)$ as possible. One sample $\mathbf{e}^T(ß)\mathbf{p}$ is too few because it could vanish though $\mathbf{i}^T\mathbf{p}$ does not; can you see how?

Two samples suffice; this is remarkable because they are too few to recover (interpolate) $\mathbf{p}$. The proof that two samples suffice comes about because $\mathbf{i}^T = \mathbf{q}_o{}^T\sqrt{2}$, which is orthogonal to $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$, so $\mathbf{i}^T\mathbf{Q} = (\sqrt{2}, 0, 0, 0)$. Another solution of this last equation turns out to be $\mathbf{i}^T = \mathbf{e}^T(-1/\sqrt{3}) + \mathbf{e}^T(1/\sqrt{3})$, as can easily be verified. The last formula is called "Gaussian Quadrature." It is a special case n = 2 of a general phenomenon; n samples artfully situated suffice to determine the integral $\mathbf{i}^T\mathbf{p}$ of any polynomial $\mathbf{p}$ of degree less than 2n .

What good comes from thinking of a cubic polynomial as a point or vector in a 4-dimensional space? There are two benefits. One is a kind of consolidation of knowledge that makes it all easier to remember; knowledge gained in one area, say calculus, seems more familiar when it is described in terms that remind us of what we learned in another area, say geometry, and *vice-versa*. But this benefit is invisible to people who do not yet know enough to be worried about how to remember it all. The second benefit is an ability to "ambush" problems, to overwhelm them by attack from unexpected directions, as will happen in other Math. courses.